Erkenn (2013) 78:153–178
DOI 10.1007/s10670-013-9507-7

ORIGINAL ARTICLE

Emergence, Closure and Inter-level Causation in Biological Systems

Matteo Mossio · Leonardo Bich · Alvaro Moreno

Received: 10 July 2013 / Accepted: 10 July 2013 / Published online: 31 July 2013

Abstract In this paper, we advocate the idea that an adequate explanation of biological systems requires appealing to organizational closure as an emergent causal regime. We first develop a theoretical justification of emergence in terms of relatedness, by arguing that configurations, because of the relatedness among their constituents, possess ontologically irreducible properties, providing them with distinctive causal powers. We then focus on those emergent causal powers exerted as constraints, and we claim that biological systems crucially differ from other natural systems in that they realize a closure of constraints, i.e. a second-order emergent regime of causation such that the constituents, each of them acting as a constraint, realize a mutual dependence among them, and are collectively able to self-maintain. Lastly, we claim that closure can be justifiably taken as an emergent regime of causation, without admitting that it inherently involves whole-parts causation, which would require to commit to stronger ontological and epistemological assumptions.

1 Introduction

Whether adequate explanations in biology require appealing to an emergent and distinctive causal regime seems to have an obvious positive answer, insofar as biological systems evolve by natural selection (Mayr 2004, p. 31). Yet, as Salmon has pointed out (1998, p. 324), one can distinguish between etiological explanations,

M. Mossio
IHPST (CNRS/Université Paris I/ENS), 13, rue du Four, 75006 Paris, France
e-mail: matteo.mossio@univ-paris1.fr

L. Bich (✉) · A. Moreno
Department of Logic and Philosophy of Science, IAS-Research Centre for Life, Mind and Society, University of the Basque Country, Avenida de Tolosa 70, 20018 Donostia-San Sebastián, Spain
e-mail: leonardo.bich@ehu.es

which tell the story leading to the occurrence of a phenomenon, and constitutive explanations, which provide a causal analysis of the phenomenon itself. Accordingly, whereas it goes without saying for etiological explanations, there seems to be no obvious answer to the question of whether a constitutive explanation of biological systems would also appeal to a distinctive regime of causation, emergent from and irreducible to that at work in physical and chemical natural systems.

During the last 40 years, the idea that the constitutive organization of biological systems does realize a distinctive regime of causation has been put forward by a number of pioneering authors like Rosen (1972, 1991), Piaget (1967), Maturana and Varela (1980), Varela et al. (1974), Varela (1979), Pattee (1972, 1973), Ganti (1975, 2003) and recently developed in various fields, including Theoretical Biology (Kauffman 2000), Biochemistry (Luisi 2006; Cornish-Bowden et al. 2007) and Systems Biology (Hofmeyr 2007). Broadly speaking, the common background assumption of this scientific trend, despite the differences among the various formulations, consists in an understanding of biological systems which translates into scientific terms an idea originally exposed by Immanuel Kant in his *Critique of Judgement*, according to which a biological system can be conceived as a *natural purpose*, such that:

> The parts of it produce themselves together, one from the other, in their form as much as in their binding, reciprocally, and from this causation on, produce a whole. In such a product of nature each part, at the same time as it exists throughout all the others, is thought as existing with respect to the other parts and the whole, namely as instrument (organ). It is then - and for this sole reason - that such a product, as organized and organizing itself, can be called a natural purpose (Kant 1987, §65, p. 287).

In this view,[1] then, biological systems realize a distinctive causal regime in which a set of constituents produce and maintain each other through a network of mutual interactions, such that the whole system can be said to be collectively able to self-produce and self-maintain. In the literature, this regime is usually referred to as *closure* (see also Mossio and Moreno 2010, for a recent analysis).[2]

In spite of the increasing scientific and philosophical work on this notion, however, the status of closure as a distinctively biological causal regime has not yet been assessed, we hold, in adequate terms. Indeed, the very idea that closure would be 'distinctive' of the biological domain seems to require adopting a non-reductionist stance, according to which biological systems realize a regime of causation that is irreducible to those at work in other classes of natural (i.e. physical and chemical) systems. In this sense, hence, the philosophical discussion on closure is related to that on emergence, to the extent that closure can be 'irreducibly

[1] See also Weber and Varela (2002) for a discussion of the Kantian rooting of this scientific tradition.

[2] It is worth noting that this meaning of 'closure' has nothing to do with Kim's one, which is at work in his argument about the "causal closure of the physical domain". According to the latter, as Kim explains, "*any physical event that has a cause at time t has a physical cause at t.* This is the assumption that if we trace the causal ancestry of a physical event, we need never go outside the physical domain"(Kim 1993: 280).

biological' only if biological systems can be shown to possess emergent causal powers. Still, existing accounts do not provide clear arguments supporting the claim that closure can be taken as an emergent causal regime.

Similarly, it is at present unclear whether or not closure involves inter-level causation. At first sight it seems obvious that closure inherently relies on the causal interplay between entities at different levels of description: the integrated activity of lower-level constituents contributes to generate the higher-level organization, and the higher-level organization plays a crucial role in maintaining and regenerating its own constituents, as well as controlling and regulating their behaviour and interactions. However, the appeal to inter-level causation in biological systems may oscillate between two different interpretations of the concept: on the one hand, the causal influence of an entity located at a given level of description on an external entity located at another (upper or lower) level of description; on the other hand, the causal influence of an entity, taken as a whole, on its *own* parts. As we will discuss, while it might seem quite obvious that closure involves inter-level causation in the first sense, a more difficult issue is whether this is also true for the second sense, which requires complying with more restrictive conceptual conditions.

In this paper, we will advocate the view according to which an adequate constitutive explanation of biological systems requires appealing to the idea of organizational closure as an emergent regime of causation.

Our argument will be twofold. On the one hand, we will develop a theoretical justification of emergent causation, by arguing that the idea of emergent causal powers can be consistently understood in terms of *constraints.* We will then suggest, by providing a general characterisation, that organizational closure should be conceptualized as a network of constraints mutually interacting in a distinctive way. On the other hand, we will maintain that although the mutual relations between constraints are such that the very existence of each of them depends on their being involved in the whole organization, an emergent closed organization does not necessarily imply inter-level causation, be it upward or downward, in the restrictive sense of a causal relation between the whole and its own parts. Yet, as we will suggest, the appeal to inter-level causation in this sense (which is the philosophically more interesting and discussed one) may possibly be relevant for organizational closure, if the adequate conceptual justification were provided.

The structure of the paper is as follows. In Sect. 2 we discuss one of the main philosophical challenges to the idea of emergence—Kim's exclusion argument—by focusing on the fact that it applies to a specific account of emergence formulated in terms of supervenience and irreducibility. In Sect. 3, we recall the distinction between two dimensions of the debate about emergence—ontological irreducibility and epistemological non-derivability, and clarify that a pertinent defence against the exclusion argument can be expressed, as we will do, exclusively in terms of irreducibility. Section 4 offers a conceptual justification of emergent properties, and argues that, because of the relatedness among their constituents, configurations possess ontologically irreducible properties, providing them with distinctive causal powers. In Sect. 5, we focus on the specific case in which configurations exert distinctive causal powers as constraints acting on underlying physicochemical changes, and we provide a precise characterization of their irreducible properties

with respect to thermodynamic flow. Section 6 describes how constraints are distinctively at work in natural self-maintaining systems, and specifically in biological systems, in which they realize the causal regime labelled 'closure'. In particular, we argue that closure can be taken as a second-order emergent configuration, ontologically irreducible and provided with distinctive causal powers, including the generation of functions. Section 7 concludes the analysis, and argues that closure can be justifiably taken as an emergent biological causal regime without admitting that it inherently involves inter-level causation. Yet, the connection between closure and inter-level causation remains an open issue requiring further philosophical and theoretical investigations.

2 The Philosophical Challenge to Emergence

The very idea of some distinctively biological regime of causation cannot be justified unless it can be shown that, in some way, a given entity possesses characteristic properties by virtue of which it can exert emergent causal powers. A conceptual justification of emergence seems then to be a necessary requirement for a coherent account of biological causation.

Philosophical work on emergence began during the late-nineteenth and early-twentieth centuries with the writings of the so-called 'British Emergentists' (Mill 1843; Alexander 1920; Lloyd Morgan 1923; Broad 1925), and has developed considerably over recent decades. As has often been underscored, a central contribution to this debate was made by Jaegwon Kim, who developed one of the most articulated conceptual challenges to the idea of emergence (Kim 1993; 1997; 1998; 2006).

In a recent survey of these issues (Kim 2006), Kim recalls what are, in his view, the two necessary ingredients of the idea of emergence, i.e. supervenience and irreducibility. *Supervenience* is a relation by virtue of which the emergent property of a whole is determined by the properties of, and relations between, its realizers. As he puts it (Kim 2006, p. 550):

> *Supervenience*: If property M emerges from properties N_1, …, N_n, then M supervenes on N_1, …, N_n. That is to say, systems that are alike in respect of basal conditions, N_1, …, N_n must be alike in respect of their emergent properties.

In turn, *irreducibility*, and more precisely, according to Kim, *functional* irreducibility, is expressed as follows:

> *Irreducibility of emergents*: Property M is emergent from a set of properties, N_1, …, N_n, only if M is not functionally reducible with the set of the N_s as its realizer (Kim 2006, p. 555).

Given the account of emergent properties in terms of supervenience and irreducibility, the central issue is whether these properties can possess distinctive causal powers. In his work, Kim has developed several lines of criticisms vis-à-vis emergence. The one, which is specifically relevant here, claims that emergent properties are exposed to the threat of epiphenomenalism. Kim's argument on this matter is known as the 'exclusion

argument', and it has been offered on several occasions. Very briefly, the idea is the following. If an emergent property M emerges from some basal conditions P, and M is said to cause some effect, one may ask "why cannot P displace M as a cause of any putative effect of M?" (Kim 2006, p. 558). If M is nomologically sufficient for whatever effect X, and P is nomologically sufficient for M (because of the supervenience relation), it seems to follow that P is nomologically sufficient for X, and M is "otiose and dispensable as a cause" of X. As a result, invoking the causal power of emergent structures would be useless, since epiphenomenal.

The exclusion argument has crucial implications for the debate about emergence and reduction. If one admits (1) that the relation between M and P is correctly described in terms of supervenience and (2) the validity of what we will call here the *principle of the inclusivity of levels*,[3] i.e. "the idea that higher levels are based on certain complicated subsets from the lower levels and do not violate lower level laws" (Emmeche et al. 2000, p. 19), then two problematic consequences follow.

First, explanation is exposed to the danger of causal drainage. Indeed, if the causal powers of an emergent entity can be reduced to the causal powers of its constituents, and if, as it may indeed be the case, there is no 'rock-bottom' level of reality, then it seems that "causal powers would drain away into a bottomless pit, and there would not be any causation anywhere" (Campbell and Bickhard 2011, p. 14).[4] Second, if there were some scientifically justifiable rock-bottom level of reality (which is a far from trivial assumption),[5] and causal drainage were blocked, the exclusion argument would force reductive physicalism (see Vicente 2011 for a recent analysis). In this second case, whatever appeal to distinctively biological causal relations (such as closure itself, and related notions such as 'integration', 'control', 'regulation', …) would at best constitute heuristic tools, unless it could be demonstrated that they can adequately be reduced to physical causation or, more generally, to any 'more fundamental' regime of causation.

[3] We take here the notion of "inclusivity of levels" as analogous to Kim's "Causal Inheritance Principle" (Kim 1993, p. 326), according to which if a property M is realized when its physical realization bases P is instantiated, the causal powers of M are identical with the causal powers of P. By the choice of "inclusivity of levels" we want to emphasize the idea that in the natural world all causes are physical or are the result of the interaction between physical entities: no special causes (vitalist, spiritual, etc., that are not physically instantiated) are introduced at different levels, e.g. at the biological and the mental ones. It should be noted that Kim's argument also requires the Causal Closure Principle as a premise, in the sense that the ultimate reduction of an emergent property to its fundamental realization base is possible only if the basal level is causally closed (Kim 2003). Yet, we maintain that the validity of the inclusivity of levels does not necessarily require appealing to causal closure: emergent causal powers can be reduced to basal powers even though the latter are not shown or supposed to be closed. Consequently, the argument we develop in this paper does not depend on the Causal Closure Principle.

[4] For Kim's purposes, the exclusion argument is originally targeted at mental causation and is not supposed to imply causal drainage. As a matter of fact, Kim himself has vehemently tried to avoid causal drain drainage as the ultimate consequence of the argument of this argument in favour of reduction. In addition, on the basis of a commitment to the Standard Model and its bottom level of fundamental physical particles, he rejects the arguments based on the possibility of the absence of a rock-bottom level of reality. For a detailed discussion of these issues see, for example Block's criticism of Kim's reduction argument (Block 2003) and Kim's reply (Kim 2003).

[5] The idea of a basic level with self-sufficient basic entities has been deeply questioned in microphysics, the very domain reductionist approaches appeal to as fundamental, where relational and heuristic accounts have taken place (Bitbol 2007).

In both cases, the very possibility of biological explanation is menaced. An adequate justification of a distinctive regime of biological causation should be provided in order to (1) avoid the danger of endless causal drainage and (2) make biological explanation theoretically independent from the physical and chemical ones, and directly related to the specificity of biological phenomenology instead of being derived from lower level explanations, and dependent on a single physical 'theory of everything' (Laughlin and Pines, 2000).

3 Irreducibility Versus Non-derivability

Before addressing the exclusion argument, let us make a preliminary conceptual distinction between two dimensions of the debate about emergence, i.e. irreducibility and non-derivability.

The exclusion argument challenges the status of emergent properties as causal agents of the world: how can a property be supervenient on something while being, at the same time, irreducible, and then possessing distinctive causal powers? An appropriate reply should then deal with the ontological issue of irreducibility, and justify emergent properties by showing that they are something ontologically 'new' with respect to their realizers. Irreducibility, hence, is inherently linked to *ontological novelty*.

Irreducibility should not be confused with the possible non-derivability of emergent properties from the emergence base, which is an *epistemological* issue. Non-derivability refers to the fact that given a description of the properties of the realizers, it is not possible to predict, explain or deduce the emergent properties of the whole.

As a matter of fact, most of the philosophical debate has tended to merge the two issues,[6] and to take both irreducibility and non-derivability as marks of emergence: emergent properties are not only irreducible but also, and crucially, non-derivable. Consider for instance the classical distinction between 'resultant' and 'emergent' properties, which is based precisely on criteria of non-derivability (or "non-deducibility", in Kim, 2006, p. 552).[7] Resultant properties are aggregative properties, which the whole possesses at *values* that the parts do not (i.e. a kilogram of sand has a mass that none of its constituents has). Emergent properties, in turn, are properties of a *kind* that only the whole possesses, whereas the parts do not (i.e. a system can be alive, whereas none of its parts is alive). Although resultant properties can be said to be, in a general sense, irreducible to the properties of their realizers, however, they are not what British emergentists (and most contemporary authors) had in mind when speaking about emergence. In fact, when appealing to

[6] The distinction has however been formulated, for instance by Silberstein and McGeever (1999), according to which *epistemological* emergence concerns models or formalisms, while *ontological* emergence involves irreducible causal capacities. Here, we follow this conventional distinction.

[7] Van Gulick (2001) refers to resultant and emergent properties as "specific value emergent" and "modest kind emergent" properties, respectively.

notions like 'unpredictability' or 'unexplainability' as the mark of emergence, most authors are focusing on epistemological non-derivability.[8]

Yet, we maintain that ontological irreducibility and epistemological non-derivability are logically distinct dimensions, and call for independent philosophical examinations. In what follows, we will discuss them separately, as two different issues.

On the one hand, we will develop, throughout most of the paper, a philosophical defence of emergence against the exclusion argument and the danger of epiphenomenalism, by relying exclusively on the irreducibility of emergent properties, without addressing the issue of their non-derivability. Emergent properties, we will argue, can be defined *exclusively in terms of irreducibility* and, crucially, they provide the system with distinctive causal powers *even though* they are derivable from their emergence base.

On the other hand, the issue of the non-derivability of emergents may play an important role in the discussion on whether emerging properties enable a system to exert inter-level causation between the whole and the parts. As we will suggest in the last part of the paper (Sect. 7.2), if an emergent property is proven to be also non-derivable from the properties of the constituents, because of the epistemological gap between them, it may be possible to interpret the relation between the whole and the parts as involving inter-level causation.

4 Irreducibility and Emergence

The aim of this section is to offer, in response to the exclusion argument, a conceptual justification of emergent properties provided with irreducible and distinctive causal powers. The core of the argument consists in suggesting that a coherent account of emergent causal powers can be obtained by rejecting the identification between the 'supervenience base' and the 'emergence base' of a property. As we will propose, a property of a whole can be functionally reducible to the set of properties of its constituents (its supervenience base) while being functionally irreducible to, and then emergent on, various categories of entities which are distinct from that set. We will argue that, once the distinction between the supervenience and emergence base is conceded, the resulting account of emergence eludes the exclusion argument and justifies the existence of distinct regimes of causation, even by maintaining the principle of the inclusivity of levels.

The argument will proceed in two steps. First, we will advocate (Sect. 4.1) an interpretation of the relation between the whole and the parts in terms of relational mereological supervenience, according to which a supervenience relation holds between the whole and the *configuration* of its own constituents, and not the collection of constituents taken separately. We will then put forward a *constitutive* interpretation of relational supervenience, according to which supervenient properties can be in principle reduced to the configurational properties of the

[8] Crutchfield, for instance, distinguished two different definitions and classes of models of emergence according to two different limitations in our capability "in principle" to describe emergent phenomena: nonpredictability and nondeducibility (Crutchfield 1994).

supervenience base. The main implication is that a supervenient property M and its basal properties S_1, …, S_n have identical causal powers. In the adoption of such a constitutive interpretation of relational supervenience lies the *monistic* stance of our perspective.

Second, we suggest (Sect. 4.2) that, even under the constitutive and monistic interpretation of relational mereological supervenience, a relation of emergence (as irreducibility) holds not between M and configurational properties but, instead, between configurational properties and the properties of different categories of entities which do not belong to the configuration. As a consequence, configurations can be justifiably said to possess irreducible and emergent properties and hence be able to exert non-epiphenomenal causal powers (in particular, as recalled in Sect. 5, as *constraints*).

4.1 Supervenience and Constitution

The logic of the exclusion argument is based on the way in which the relation between an emergent property M of the whole W and the set of basal properties N_1, …, N_n of its constituents P is conceived. Namely, the relation is supposed to be simultaneously one of (mereological) supervenience and functional irreducibility, while assuming at the same time, as mentioned above, the validity of the principle of inclusivity of levels.

In *Mind in a Physical World*, Kim paved the way for an answer to the exclusion argument able to maintain the inclusivity of levels, by clarifying the terms of the supervenience relation, and particularly specifying how the supervenience base is to be conceived. Kim argues that emergent properties are *micro*-based *macro* properties, i.e. second order properties emerging out the first-order properties and relations of the basal constituents (Kim, 1998, pp. 85–86). The central idea is that the relevant supervenience base is not a set of properties of constituents taken individually or as a collection, but rather the properties of the *configuration* of constituents, i.e. the whole set of inherent *and* relational properties of the constituents. In other words, mereological supervenience should not be interpreted as atomistic but, rather, relational (see also Thompson 2007, pp. 427–428).[9]

The move to adopt relational mereological supervenience makes configurations of constituents the relevant supervenience base. The basal properties S_1, …, S_n that bring about a supervenient property M are not the properties of the collection of constituents taken separately, but rather the configurational properties of the

[9] The debate between a relational and an atomistic interpretation of the supervenience and emergence base has a long history that dates back to the first formulations of the notion of emergence in the British Emergentism. In Alexander's framework, for example, space and time, the lower level on which the whole natural world emerge, are relational concepts, not definable separately (Alexander 1920). The opposition between atomistic and relational approaches is particularly evident in Lloyd Morgan's work. In contrast to the billiard balls model of extrinsic interactions, he presents the idea of relatedness based on inherent relations, that contributes to specify the properties of the terms involved in the relation (Lloyd Morgan 1923, p. 19). It is also worth noting that according to some authors, Kim's reference to relations is still made in a fundamentally atomistic framework, and does not imply a clear commitment to relational mereological supervenience, which implies the idea that relations "do not simply influence the parts, but supersede or subsume their independent existence in an irreducibly relational structure" (Thompson 2007, p. 428).

constituents *qua* constituents (including their mutual *relations*, which alter their intrinsic properties as separate elements), which appear only when the configuration is actually realized. If the basal constituents actually and collectively constitute a global pattern or system W, then their properties would now include those generated by their being involved in specific relations and interactions with others elements.

The adoption of relational mereological supervenience has relevant implications for the question concerning the distinctive causal powers of the supervenient property with respect to its supervenience base. Indeed, the idea that emergent properties would be reducible to the properties of the constituents taken in isolation seems to be excessively committed to an atomistic view of nature, which does not take relations into account (Campbell and Bickhard 2011). In turn, the claim that a supervenient property M is in principle reducible to the set of configurational (i.e. including relations) properties S_1, …, S_n of its constituents is more convincing (again, by assuming the principle of inclusivity of levels), since configurations are far richer and more complex determinations than the mere collection of intrinsic properties of constituents.

Accordingly, we hold that relational supervenience does not imply functional irreducibility but, on the contrary, *constitution*: M supervenes on S_1, …, S_n since it consists in S_1, …, S_n. A supervenient property M of a whole W corresponds to the whole set of configurational properties S_1, …, S_n of its constituents (its supervenience base B). The set of the (relevant) configurational properties of its constituents of the system is, at least in principle, equivalent to the supervenient property. Hence, if M can be functionally reduced to the set S_1, …, S_n of configurational properties of its constituents, it follows that it cannot possess distinctive causal powers[10] since, in fact, they are equivalent.[11]

Yet, as we will claim in the following section, a coherent account of emergent properties provided with distinctive causal powers can still be provided, even under the constitutive interpretation of whole-parts relations.

4.2 A Reply to the Exclusion Argument

Our reply to the exclusion argument consists in arguing that even though supervenient properties (M) have no distinctive causal powers with respect to the configurational properties S_1, …, S_n of the constituents, S_1, …, S_n themselves (which are equivalent to M, because of constitution) are irreducible properties which may generate distinctive causal powers. Accordingly, S_1, …, S_n can be said to be genuinely emergent. In other terms, there is an interpretation of emergence which is compatible with a monistic stance.

Here is our argument. A given configuration C of elements of a whole W is identified by a set of (possibly dynamic) distinctive constitutive and relational

[10] The interpretation of relational mereological supervenience in terms of constitution is consistent, we argue, with the position developed by Craver and Bechtel (2007) within their mechanistic framework. As they suggest, the relations between constituents located at different levels in a mechanism are better understood as constitutive relations (pp. 554–555). See Sect. 7 below for a detailed discussion.

[11] For simplicity, we will only refer, from now on, to 'configurational properties S_1, …, S_n' (equivalent to 'property M'), and to 'whole W' (equivalent to 'supervenience base B').

properties S_1, …, S_n. On the basis of this set of distinctive properties, a configuration is functionally irreducible to whatever entity that does not *actually*[12] possess the same set of properties. We claim that a relation of emergence holds between a configuration C and any emergence base P whenever C is irreducible to P, i.e. if C possesses some distinctive set of configurational properties that P does not possess, such that C does *not* supervene on P. The reader would immediately note that this characterization of emergence is very general, and could in principle include a wide range of obvious and uninteresting cases of P, which would not be considered salient for the philosophical debate on emergence. This is correct, and we deal with this issue just below. Yet, let us point out here that, as Campbell and Bickhard (2011, p. 18); see also Teller 1986) have highlighted, appealing to configurations seems to be a sufficient answer to the danger of causal drain and epiphenomenalism. The crucial point, as we mentioned above, is that configurations include relational properties, which cannot be reduced to intrinsic properties, i.e. properties of constituents taken in isolation. Relatedness is ontological novelty. As a consequence, because of relatedness (again: *actual* relatedness), configurations may possess distinct causal powers that would not exist otherwise.

To avoid confusion, it is important to stress again that this account, in contrast with most existing ones, defines emergence exclusively in terms of ontological irreducibility, by leaving aside the issue of the epistemological non-derivability of C from P. C is emergent on P if it possesses some set of *new* (relational) properties S_1, …, S_n which P does not possess, and which are then irreducible to the set of property N_1, …, N_n of P. A distinct issue, which is irrelevant here, is whether one can derive or predict S_1, …, S_n from N_1, …, N_n. In particular, S_1, …, S_n would be irreducible *even if* they were derivable, because of the novelty introduced by the relations among constituents.

At this point, given the constitutive relation between the whole and the constituents advocated so far, one may wonder what exactly configurations emerge on. Following our definition, three main kinds of emergent base P can be logically identified. First, the configuration C is not supervenient, yet is emergent on the properties of any proper *subset* P_{sset} of its constituents (its parts). A wheel has emergent properties and distinctive causal powers on any subset of itself (i.e., a half-wheel). Second, the configuration C is not supervenient, yet is emergent on its *substrate* P_{sstr}, i.e. the collection of its constituents taken separately as if they were not constituents (so to speak, the 'potential ingredients' of a configuration). A wheel is emergent on the collection of molecules taken as if they were not actually assembled as a wheel.

Third, and importantly, the configuration C is not supervenient, yet is emergent on its *surroundings* P_{surr}, i.e. each set of external elements that does not actually constitute C. The wheel is emergent on each set of external molecules or entities, which are not actual constituents of it. In particular, given that a very broad set of entities might be included in P_{surr}, only relevant instances will actually be

[12] It is important to emphasise that configurational properties must be actually realised, and not just "dispositional". As a consequence, a configuration C is functionally irreducible, in this account, also to those entities that would possess the "potential disposition" to actualise these properties.

considered: in particular, the reference to surroundings P_{surr} will be restricted to those relevant P_{surr} on which the configuration C has causal effects, by virtue of its emerging properties. As we will discuss in the following section, this is precisely the relevant case with regard to biological systems.

At this point, we have all the elements required to formulate our reply to the exclusion argument. The argument claims that properties cannot be emergent unless it can be shown that they possess distinctive causal powers; at the same time, it seems that, as supervenient properties, they do not possess new causal powers with respect to their supervenience base. Hence, they are epiphenomenal. To this, we reply that emergent properties do not need to be irreducible to their supervenience base to possess distinctive causal powers: what matters is that configurations, because of relatedness, possess irreducible properties with respect to their subsets, substrate and (relevant) surroundings. Supervenience and emergence are then *alternative* notions: either a set of properties is supervenient on another one (in which case there is constitutive relation between them), or it is emergent (in which case there is irreducibility).

Let us stress again that this way of conceiving emergence, interpreted exclusively as ontological irreducibility, is indeed very general. For instance, all chemical bonds are configurations emergent on their parts, substrate and surroundings, since they realize new relations, and therefore possess distinctive configurational properties. Yet, the fact that this definition covers also irrelevant or obvious cases is, we argue, the price to pay for making it compatible with the constitutive interpretation of the relations between the whole and the parts. More generally, we hold that this characterisation of emergence is sufficient to provide a justification for the appeal to distinctive and irreducible causal powers in the scientific discourse (Laughlin et al. 2000), and specifically in biology. Emergence appears whenever scientists are dealing with a system, as biological ones, whose properties are irreducible to those of its isolated parts, substrate and surroundings. In such cases, one must introduce new objects, relations and causal powers, which exist only within that very system, and not in its emergent base.[13]

5 Emerging Causal Powers as Constraints

By virtue of their relatedness, configurations possess emergent properties, and may exert distinctive causal powers on their surroundings that can take different forms, following the kind of systems under consideration. Let us focus here on the case in which these causal powers are exerted as *constraints*, which can be organized in turn as closure.

In general terms, constraints are determinations that reduce the degrees of freedom of the system to which they belong (i.e. an inclined plane which reduces to two spatial dimensions the motion of a ball on it), simplify (or change) the

[13] It is worth noting that the relation between the emergent properties and its emergence base can be interpreted both synchronically and diachronically. Being based on novelty, in fact, the irreducibility to any entity that does not belong to an actual configuration is in principle compatible with both the dimensions of emergence. See also footnote 21 below.

description of that system, and contribute to provide an adequate explanation of its behaviour, which might otherwise be under-determined or wrongly determined (Umerez and Mossio 2013).

In a given system, constraints are characterized as configurations emerging on, and acting on, specific surroundings P_{surr}, i.e. a set of physicochemical changes that involve the movement, alteration, consumption, and/or production of entities in conditions away from thermodynamic equilibrium. More precisely, given a particular process P_{surr}, a configuration C acts as a constraint if the following conditions are fulfilled:

1. At a time scale τ_i, *C is conserved throughout* P_{surr}, i.e. there is a set of emerging properties S_1, …, S_n of C which remain unaffected throughout P_{surr};
2. At τ_i, *C exerts a causal role on* P_{surr}, i.e. there is some observable difference between P_{surr} and P^c_{surr} (P^c_{surr} is P_{surr} under the causal influence of C by virtue of properties S_1, …, S_n);

It is worth noting that, usually, *C does not extend the set of possible behaviours of* P_{surr}, i.e. P_{surr} could in principle (although very unlikely) exhibit, at different time scales, the behaviour of P^c_{surr} without the action of C.

Consider the example of the vascular system, taken as a global constraint (C) on the flow of oxygen (P_{surr}) in an organism. First, the topology of the vascular system remains unaltered during the oxygen flow (at least at relatively short scales τ_i). Second, the vascular system exerts a causal role on the flow of oxygen, since, for instance, the flow takes the form of a transport (canalized) to the neighbourhood of each cell, instead of having a diffusive (unconstrained) form. In this case, moreover, the flow of oxygen could reach (at least in principle, since in practice this would be extremely unlikely) each cell at an adequate rate even in the absence of the vascular system, from the point of view of statistical mechanics.

For the purposes of this paper, two features of the concept of constraint have to be emphasized.

First, insofar as they are causal effects produced by emerging properties of configurations, constraints depend on the existence of the relevant configuration. Hence, constraints are not universal, but rather local and contingent causes. In particular, configurations fit the definition only at some specific time scales, whereas at different scales they usually fail to satisfy one or more conditions. For instance, at time scales larger than τ_i, configurations are normally subject to degradation, and must be replaced or repaired (and then fail fitting condition 2). Yet, at τ_i, constraints are irreducible to the thermodynamic flow (in particular because of their conservation) and require an alternative description (Pattee 1972). Because of their distinctive causal powers, which cannot be reduced to those of their emergent base P, constraints constitute an *emergent regime of causation*, operating in addition to physicochemical changes.

Second, in the description of physical systems, constraints are in most cases introduced as external determinations (boundary conditions, parameters …), such that they exert some causal influence on P_{surr}, while their existence (or, more precisely: the maintenance of the specific properties S_1, …, S_n providing them with relevant causal powers) does not depend on P_{surr}. For instance, the inclined plane

constrains the dynamics of the ball, but the constrained dynamics do not have a causal role on the existence or maintenance of the plane.

The appeal to constraints to provide adequate descriptions and explanations is ubiquitous in natural sciences. In the biological domain, however, constraints play a characteristic role to the extent that they give rise to a specific causal regime of self-maintenance, that we label *closure*.

6 Self-maintenance: From Self-organisation to Closure

In many natural systems, the relation between a constraint C and its emergent base P_{surr} is oriented, in the sense that C may causally act on P_{surr}, but not vice versa. Yet, there are cases in which C acts on P_{surr} that, once constrained, contributes to determine (at least some of) the boundary conditions at which C exists. To use our labels, there are cases in which C constrains P_{surr} such that the constrained P^{c}_{surr} constitute some of the boundary conditions required for C to exist. In that case, the resulting system realizes *self-maintenance*, since C constrains P_{surr}, which in turn, once constrained, maintains C. If C did not act on P_{surr}, it would not (or it would cease to) exist. Self-maintaining systems are then systems able to maintain some of the conditions required for their own existence, by virtue of the constraining action of their own configurations (Mossio and Moreno 2010, p. 272; see also Kauffman 2000, for a related analysis in terms of 'work-constraint cycles').

Self-maintenance exists in the physical and chemical domain. The classical example are dissipative structures (Glansdorff and Prigogine 1971; Nicolis and Prigogine 1977), in which a huge number of microscopic elements adopt a global, macroscopic ordered configuration (a 'structure') in the presence of a specific flow of energy and matter in far-from-thermodynamic equilibrium (FFE) conditions. In turn, the macroscopic configuration exerts a constraint that contributes to the maintenance of the FFE flow of energy and matter enabling the persistence of the microscopic dynamics (Ruiz-Mirazo 2001, p. 59).

A number of physical and chemical systems, such as Bénard cells, flames, hurricanes, and oscillatory chemical reactions, can be pertinently described as self-maintaining dissipative systems. Let's take the example of 'Bénard cells', i.e. macroscopic structures that appear spontaneously in a liquid when heat is applied from below (Chandrasekhar 1961). In the initial situation, in which there is no difference in temperature between upper and lower layers, the liquid appears uniform, in terms of the statistical distribution of the kinetic energy of the molecules. When heat is applied, and the temperature in the lower layer is increased up to a specific threshold, the liquid's dynamics change dramatically: the random movements of the microscopic molecules spontaneously become ordered, creating a macroscopic pattern (convection cells). In each cell, billions of microscopic molecules rotate in a coherent manner along a hexagonal path, either clockwise or anticlockwise, and always in the opposite direction from that of its immediate neighbours in a horizontal plane.

Bénard cells appear when P_{surr} realizes some specific boundary conditions (e.g. the heat applied from below), which exert external constraints on the dynamics. Yet,

once they have appeared, the maintenance of Bénard cells depends not only on those independent boundary conditions, but also on some constraint exerted by the very configuration on its surroundings. For instance, the cells *capture* surrounding water molecules in their dynamics, turning them into constituents. It is through this action that Bénard cells contribute to maintain some of the boundary condition at which the flow of energy and matter traversing them may occur. With our labels, the emerging configuration C acts as a constraint on P_{surr}, on which it emerges, by turning them into appropriate boundary conditions P^{c}_{surr}. We need then to appeal to C itself to explain its own maintenance, which would otherwise be impossible (or at least very unlikely) on the basis of the unconstrained properties of P_{surr}.

As it has been recently emphasized (Mossio and Moreno 2010), dissipative systems realize a *minimal* form of self-maintenance, in the sense that they generate a *single* macroscopic structure acting as a constraint on its surroundings. Accordingly, dissipative systems make a unique contribution to their own maintenance, since they contribute to maintain the unique constraint involved in the self-maintaining loop between C and P_{surr}.

When considering biological systems, the situation is more complex. Unlike minimal self-maintaining systems, biological systems generate a network of structures, exerting mutual constraining actions on their boundary conditions, such that the whole organization of constraints realizes *collective* self-maintenance. In biological systems, constraints are not able to achieve self-maintenance individually or locally: each of them exists insofar as it contributes to maintain the whole organization of constraints that, in turn, maintains (at least some of) its own boundary conditions. Such mutual dependence between a set of constraints is what we call *closure*, the causal regime that, we claim, is paradigmatically at work in biological systems.[14]

Let's give a more formal characterization. A system realizes closure if and only if it contains a set of structures C_1, …, C_n acting as constraints at the relevant time scales τ_1, …, τ_n, by virtue of emerging properties S_1, …, S_n such that, for each constraint C_i (at least some of) the boundary conditions required for the maintenance of the relevant emergent property M_i are determined by the causal action of another constraint C_j by virtue of a property M_j, whose maintenance depends in turn on the action of C_i.

As all dissipative systems, be they physical or chemical, biological systems are traversed by a flow of energy and matter, which takes the form of processes and reactions occurring in open thermodynamic conditions. What specifically characterizes biological systems[15] is the fact that the thermodynamic flow is constrained and canalized by a set of constraints in such a way that they realize a mutual dependence. In this sense, biological organization is a specific kind of *higher-level* configuration, which achieves self-determination as collective self-constriction: the conditions of existence of (at least a subset of) the constitutive constraints are, because of closure,

[14] The concept of closure has been proposed by several influential authors, that we have mentioned in the introduction. For a recent survey, see Chandler & Van De Vijver (2000). Mossio (2013) provides a synthetic overview of the meaning and uses of the term in the biological domain.

[15] Or, at least, systems being 'at the edge' of the biological domain. We do not discuss this issue here, since it does not interfere with the central point.

mutually determined within the very organization. In what follows, the term 'organization' will then specifically refer to this kind of higher-level configurations, whose constituents are themselves configurations, each of them acting as constraints.

This characterization is, of course, very general and lacks several specifications about the way in which closure is actually realized in biological systems. Yet, it is precise enough for the purposes of this paper, and in particular, allows discussing three crucial implications.

First, the closed regime of dependence among constraints is emergent on, and irreducible to, the open regime of thermodynamic processes and changes. Closure is realized through the mutual causal action of the constraints which, as discussed in Sect. 5, are irreducible to the thermodynamic flow, each constraint being conserved at the relevant time scale τ_1, …, τ_n. As a consequence, a reductive description of closure in terms of the causal regime of thermodynamic changes would be inadequate, since it would be unable to include constraints and their contribution as causal factors.[16] In particular, a description of biological organization not appealing to the causal power of constraints and their closure would amount to a system constituted by a cluster of unconnected processes and reactions, whose coordinated occurrence could be theoretically possible at very large time scales (see the definition of constraints above), but extremely unlikely—and often impossible—at biologically relevant time scales.[17]

Second, organizations themselves possess, because of closure, emergent properties. One of them, which is particularly relevant here, is the generation of functions. As it has been recently argued (Mossio et al. 2009; Saborido et al. 2011), when subject to closure, constraints correspond to biological functions: performing a function, in this view, is equivalent to exerting a constraining action on an underlying process or reaction.[18] All kinds of biological structures and traits to which functions can be ascribed satisfy the above definition of constraint, although at very different temporal and spatial scales. Some intuitive examples, in addition to the vascular system mentioned above, include, at different scales: enzymes (which constrain reactions), membrane pumps and channels (which constrain the flow of ions through the membrane), organs (as the heart which constrains the flow of blood), and so on. The emergence of closure is then the emergence of functionality

[16] It is of course conceivable that a description of constraints might possibly be given in terms of thermodynamics, specifically as entities *not being affected* by the thermodynamic flow. However, in this case, constraints (and then closure) would not be reduced to a different causal regime, but simply redescribed in different terms.

[17] This allows distinguishing, moreover, a closure of constraints from a cycle of processes or reactions as, for instance, the water cycle. In the case of cycles, the entities and processes involved in the cycle (e.g. clouds, rain, spring, river, sea, clouds …) do not act as constraints on each other, and the system can be adequately described by appealing to a set of external boundary conditions (ground, sun …) acting on a single causal regime of thermodynamic changes (see also Mossio and Moreno 2010).

[18] In this framework, functions are interpreted in an organizational sense: a trait is functional if and only if it exerts a constraint that is subject to closure and, consequently, contributes to the maintenance of the organization while being maintained by that very organization. As extensively discussed in Mossio et al. (2009) the organizational account of functions integrates in a unified framework both etiological and causal-systemic theories of function.

within biological organization: constraints do not exert functions when taken in isolation, but only insofar as they are subject to a closed emergent organization.

Third, closure is specifically defined with respect to the emergence base P_{surr} constituted by a set of processes and changes occurring in conditions far from thermodynamic equilibrium. Although mutually irreducible, the two causal regimes realize a two-way interaction in biological systems, to the extent that constraints act on thermodynamic processes and changes, which in turn contribute to reproducing or maintaining these constraints. Hence, it might be tempting to conclude that closure (just as any form of self-maintenance) inherently involves not just emergent causation but also inter-level causation, at work between the two causal regimes. Yet, there are several reasons to resist the temptation, at least insofar as particularly controversial kinds of inter-level causation are invoked.

7 Inter-level Causation

The issue of inter-level (be it upward or downward) causation has been, more or less explicitly, a central aspect of the debate about emergence since its beginnings,[19] since the very concept of emergence carries on the issue of the relations between properties at different levels.

In Kim's account, attributing causal powers to emergent properties necessarily implies downward causation. Let's recall his argument that, as we discussed, identifies the supervenience and the emergence bases. Let be M and M* two emerging properties, and suppose that M causes M* (a case of 'same-level' causation). As an emergent property, M* has an emergence base, say P*. Given the supervenience relation, P* is necessary and sufficient for M*: if P* is present at a given time, then M* is also present. Accordingly, it is unclear in what sense M could play a causal role in bringing about M*: given P*, its role would be useless, unless M has in fact something to do in causing P*. In other words, the same-level causation of an emergent property makes sense only if this implies the causation of the "appropriate basal conditions from which it will emerge" (Kim 2006, p. 558). As a consequence, causation produced by emergent properties seems to imply, in all cases, downward causation in the sense of a causal influence exerted by an emergent property on the basal conditions of another emergent property.

Yet, as Kim himself has argued (Kim 2010), this general form of downward causation, i.e. a causal influence exerted by an entity at a higher-level on a distinct entity located at a lower level, is indeed widespread and unproblematic. In particular, this interpretation of downward causation applies straightforwardly to

[19] According to Lloyd Morgan, "[…] when some new kind of relatedness is supervenient (say at the level of life), the way in which the physical events which are involved run their course is different in virtue of its presence-different from what it would have been if life had been absent. […]. I shall say that this new manner in which lower events happen—this touch of novelty in evolutionary advance *depends on* the new kind of relatedness" (Lloyd Morgan 1923, p. 16). According to Stephan (1992) Lloyd Morgan's passage could admit different interpretations, such as that of a logical claim about supervenience. On the contrary, McLaughlin asserts: "In Morgan one finds the notion of downward causation clearly and forcefully articulated" (McLaughlin 1992, p. 68).

self-maintenance and closure, which inherently involve, as discussed, the upward and downward causation between constraints and dynamics, each of them located at different levels of description (see also Sect. 7.2 below).

The more controversial form of downward causation would be that exerted by a whole on its own constituents (in Kim's terms, "reflexive" downward causation, 2010, p. 33). According to Emmeche et al. (2000), there are various possible interpretations of reflexive upward and downward causation. In their view, the only non-contradictory versions of the concept are those interpreting downward causation in terms of 'formal' causation (2000, pp. 31–32), such that the whole exerts *a constraining action on its own constituents*, by selecting specific behaviours among a set of possible ones. This interpretation can be taken, as Emmeche and co-authors claim, as the standard and possibly more compelling one of downward causation, and it is very close to the original proposal by Campbell (1974).[20]

As an illustration, consider Sperry's classical example of the wheel rolling downhill (Sperry 1969). On the one side, the various molecules generate the wheel as a whole. On the other side, as Emmeche et al. (2000, p. 24) explain, "none of the single molecules constituting the wheel or gravity's pull on them are sufficient to explain the rolling movement. To explain this one must recur to the higher level at which the form of the wheel becomes conceivable". The set of configurational properties of molecules is supposed here to underdetermine their behaviour so that, in order to explain it, one needs to appeal to a property of the whole (in this case: the form of the wheel) that would generate a causal influence (a selective constraint) exerted on its own constituents.

Because of the (supposed) under-determination of constituents by configurational (intrinsic and relational) properties, constituents' behaviour is partly determined, in a functionally irreducible way, by the whole to which they belong. In particular, this train of thought seems to equally apply to biological systems, in which the behaviour and dynamics of the parts appears to be, in an important sense, determined (notably through regulation and control functions) by the downward causation exerted by the whole system to which they belong.

In what follows, we will examine whether self-maintenance and closure do involve some form of reflexive inter-level causation, intended as a particular form of constraint exerted by the whole on its parts. As we will argue (7.1), there seems to be no compelling argument in favour of a positive answer in our framework, at least under the monistic assumptions adopted so far. Alternative conclusions could be obtained (7.2) by rejecting some of these assumptions, or by shifting the analysis to an epistemological or heuristic dimension.

Before continuing, a terminological clarification. A possible objection might contend that this debate somehow forces a narrow understanding of inter-level causation in terms of reflexive whole-parts causal influence, whereas the usual meaning in the biological domain refers to the non-reflexive case, where higher-level entities interact with lower-level entities, the latter not being constituents of the former. Indeed, this interpretation of inter-level causation applies straightforwardly

[20] Campbell defines downward causation as follows: "all processes at the lower level of a hierarchy are restrained by and act in conformity to the laws of the higher level" (1974, p. 180).

to biological organisation, and is inherently involved in the very notion of closure. In this sense, biological discourse requires a general concept of inter-level causation. To avoid ambiguities, we propose using different terms to refer to the two ideas: in what follows "inter-level causation" will be therefore used for the general non-reflexive case, and "nested causation" for the reflexive whole-parts case.

This way, biological descriptions would be able to refer to inter-level causation, while at the same time avoiding incongruities with philosophical analyses.

7.1 Why We Do Not Need Nested Inter-level Causation in Biology

The account of emergence and supervenience developed so far has relevant implications on the conception of nested causation.[21]

Concerning the supervenience base B—insofar as the principle of inclusivity of levels is maintained (but see Sect. 7.2 below), and the relation between an emergent property M and the configurational properties S_1, …, S_n of B is conceived as constitutive—the exclusion argument applies more cogently to relational supervenience than to its atomistic version. As a consequence, as Craver and Bechtel (2007) emphasize, no nested causation can exist between an emergent property and its own supervenience base: there is no justification for claiming either that S_1, …, S_n 'generate' or 'produce' M, or that M exerts downward causation on S_1, …, S_n. In particular, the closed organization does not exert causation on the whole network of constitutive constraints, and the whole network of constitutive constraints does not produce the closed organization. Under the monistic stance adopted so far, there is therefore no room for nested causation.

Let us consider now the emergence base P of C, and its different versions discussed in Sect. 4. Is there nested causation between the whole configuration C and any *subset* P_{sset} of its constituents? In our view, by assuming the principle of the inclusivity of levels, the answer is no, since the properties of each P_{sset} (which may refer, for instance, to each individual constraint) are by definition configurational, so that the appeal to some constraint exerted by the whole would be redundant: configurational properties are such precisely because an entity belongs to a whole. Also, no nested causation occurs between the whole and its *substrate* P_{sstr} because, in our account, the collection of its constituents taken separately (without their configurational properties) is an abstract description that does not correspond to the way in which constituents are organized in the system. Since, in the system there is no such a thing as the collection of unrelated constituents, they cannot, a fortiori, be involved in nested causation, or indeed any causation at all.

[21] In the philosophical literature, nested causation comes in two variants, synchronic and diachronic (Kim 2010, pp. 34–36). On the one hand, *synchronic* nested causation refers to the situation in which upward and downward causation would occur simultaneously. In more technical terms, a supervenient property M acts causally on its supervenience base S_1, …, S_n *at the same time* that the supervenience base generates M. On the other hand, *diachronic* (or diagonal) nested causation refers to the situation in which M acts on its own supervenience base S_1, …, S_n, causing its modification, but only at a subsequent time with respect to the upward determination. In this paper, however, we assume that the distinction is irrelevant, since we question the very idea of the causal influence of M on S_1, …, S_n, be it synchronic or diachronic.

The case of the third kind of emergence base, the *surroundings* P_{surr}, is somehow different. As we discussed in Sects. 5 and 6, emergent configurations do exert a causal action on their surroundings, notably in the form of constraints. Yet, surroundings are by definition external to the configuration, which means that the constraints exerted by C on P_{surr} *can by no means be interpreted in terms of nested downward causation*.

The claim according to which constraints, in our framework, do exert causal powers, but not in the form of nested causation, has crucial consequences on the interpretation of self-maintenance and closure.

In the case of physical self-maintaining systems, the fact that the emergent configuration acts to maintain itself does not appear to constitute, *per se*, a case of nested causation, since the constraining action is exerted on the surroundings of the configuration, not on its own constituents. Let us consider again the example of Bénard cells. An interpretation appealing to nested causation would claim that each cell (i.e. the emergent configuration) exerts a constraint on its own microscopic constituents, in the sense that the fact of belonging to a given cell *determines* whether a molecule rotates in a clockwise or counter-clockwise direction. As Juarrero (2009, p. 85) puts this: "Once each water molecule is captured in the dynamics of a rolling hexagonal Bénard cell it is no longer related to the other molecules just externally; its behaviour is contextually constrained by the global structure which it constitutes and into which it is caught up. That is, its behaviour is what it is *in virtue of the individual water molecules' participation in a global structure*".

Yet, what we call the cell is the configuration of constituents, so that, as we argued above, it is redundant to appeal to the whole set of constituents and relations to explain the behaviour of each constituent, whose characterization already includes its relational properties as part of the configuration. Once a given molecule has been 'captured' by the cell and has begun to rotate with the others, in what sense would it still be "constrained by the global structure"[22]?

Two reasons may explain why self-maintaining systems seem to be a case of nested causation. First, the description of the configurational properties of dissipative structures, which is available at a given moment, usually under-determines their behaviour. This is of course a crucial point: still, as we discussed earlier, this should not be taken as a sufficient reason to appeal to nested causal relations since, as we pointed out in Sect. 3, it confuses epistemological non-derivability with ontological irreducibility (but see Sect. 7.2 below). Second, self-maintaining systems would not exist if they did not generate a causal loop between the whole configuration and its constituents. Yet, the crucial point is that, in our view, this loop is not a *direct* loop, but rather an *indirect* one, realized through the

[22] A satisfactory analysis of downward causation, then, requires a careful distinction between two ideas. One is the idea that a configuration is made up by a set of constituents, which have causal interactions between them. Explaining why a given molecule of water is rotating in a given manner at a given moment requires an appeal to its causal interactions with other constituents. And the reason why a set of constituents may exert a causal influence on other constituents is, of course, that all of them belong to the same system. The other idea, in contrast, is that the 'whole system', including any specific constituent, would have a causal role *on that very constituent*.

action of the constraint on its surroundings. What might appear as an action exerted on the constituents is in fact exerted on the *boundary conditions* of these constituents.

In the light of these considerations, in particular, we do not think that the appeal to the supposed constraint exerted by the configuration on its own constituents in terms of *formal* causation is explanatory (again, under the monist assumptions adopted so far). The formal causation of the whole on its constituents would be in principle reducible to the constraining action exerted on the boundary conditions of these constituents, without loss of information or explanatory power.

Let us now examine closure. Is there a characteristic aspect of closure that would justify, in contrast to simple self-maintenance, the claim according to which it realizes nested causation?

The main difference between physical self-maintenance and closure is that, in the second case, self-maintenance is realized collectively, by a network of mutually dependent constraints. In real biological systems, closure is realized though a very complex organization of constraints, such that, in most cases, a given constraint exerts its action on surroundings that have already been subject to the causal influence of at least one other constraint. For instance, most enzymes act on reactions whose reactants are the result of the joint action of other constraints, including the membrane (through its channels and pumps). In these cases, it can be said that constraints act on entities which are already 'within' the system, at least in the sense of having already been constrained by the system. This seems to be a clear difference with respect to simple self-maintaining systems, and one may then conclude that the closed organization does act on its own constituents, and realizes nested causation.

Yet, we hold that the conclusion is incorrect, since it interprets those constrained processes and reactions as constituents of the organization (which, we recall, is a higher-level configuration of constraints), whereas they are not. In biological systems, the constituents of the organization are the constraints themselves, which realize collective self-maintenance. Under the constitutive interpretation of the relation between the whole and its constituents, the organization as such does not possess emergent and distinctive causal powers with respect to the network of constraints which, in turn, exerts causal powers on surroundings which are not themselves constituents of the network (although they usually are within the spatial borders of the system).[23] Accordingly, we maintain that closure does involve inter-level causation, but *not* nested causation.

A second reason why closure seems to inherently imply nested causation is that evoked by Kant (1987), i.e. the fact that the existence of the constituents (the constraints) 'depends on the whole'. Indeed, the mutual dependence among constraints is a fundamental difference between organizations and other configurations. In the second case the existence and maintenance of the constituents does not depend on their being involved in the configuration: one can decompose a wheel into its molecular elements,

[23] The physical processes on which the network exerts (constraining) causal powers can, in some cases, become members of the network itself, when they enter into configurations which act as constraints. Nonetheless, the network would exert causal powers on them as long as they are part of its surroundings, and it would cease acting causally on them as soon as they would start playing the role of constraints.

which would continue to exist as separate elements. The same holds for the microscopic constituents of a dissipative system. In contrast, closed organizations imply a more 'existential' kind of relation among constituents (the constraints themselves), which exist as far as they are involved in the whole organisation. Actually, the appeal to formal causation advocated by several authors is essentially aimed, in our view, at capturing this distinctive feature of biological organisms.

Yet, these specific features of closed organizations do not require ascribing distinctive causal powers to the whole, since closure can be realized through the network of mutual, usually hierarchical, causal interactions. 'Depending on the whole', therefore, could simply mean 'depending on the whole network of interactions' without appealing to the whole as causal agent emergent on its own supervenience base. This interpretation of the whole-parts relation in biological organization is particularly relevant because it applies to all those cases in which biological literature typically appeals to nested causation, i.e. all kinds of *regulation* and *control* mechanisms, —very common in biological systems—thanks to which organisms are able to (adaptively) compensate for internal and/or external perturbations (Piaget 1967; Fell 1997). What is frequently described as a causal action of the whole system on its own constituents, is in fact the result of the interaction among hierarchically organized constraints (or subsystems of constraints) which can result, for instance, in the acceleration of the heart rate and glucose metabolism when the organism starts playing tennis (see Craver and Bechtel 2007, p. 559, for a detailed description of this example, and other relevant ones). Regulation and control can be understood in terms of (non-nested) inter-level causal interactions among constraints: although they inherently require, as all biological functions do, the realization of closure as an emergent causal regime (see Sect. 6 above), they do not involve nested causation exerted by the whole organism.

7.2 Why We Might Need, After All, Nested Causation in Biology

The rejection of nested causation depends on the constitutive interpretation of the supervenience relation adopted so far. Indeed, the central goal of the analysis was to suggest that closure can be justifiably taken as an emergent and distinctively biological regime of causation *even* under a constitutive interpretation of supervenience. Yet, several strategies could be adopted to justify nested causation, and they might be successful and operational in some cases, including the biological domain, which is specifically under study here. To date, however, we think that these strategies lack any compelling argument in favour of their adoption in the biological domain; their relevance is still under conceptual and scientific scrutiny. That is why, in our view, the constitutive interpretation of the whole-parts relation is still the wiser one. Let us discuss these strategies.

The first strategy is ontological and advocates that a non-constitutive interpretation of relational supervenience should be adopted, in order to admit causation of the whole on the constituents. In this interpretation, emergent properties can be at the same time supervenient on *and* irreducible to configurations. For this ontological stance to be coherent, one must accept the violation of the inclusivity of levels, hence accepting the idea that the very same entity (say: a constituent of a

configuration) may possess different properties, and therefore obey different laws or principles, at different levels of description. In other terms, it consists in rejecting the monistic stance advocated so far. For instance, each molecule constituting the wheel would have the property of behaving in a given way when considering the whole configuration, but each of them would *not* possess the same property when looked at individually. Even though we are looking at the very same molecules under the very same conditions, their properties would vary according to the level of description, since the relevant laws and principle would vary.

In our view, rejecting the principle of the inclusivity of levels could be indeed an important tool to adequately account for natural phenomena that would require appealing to nested causation. We have no principled objections to this position. Yet, we maintain that its relevance for the biological domain is still uncertain. As Craver and Bechtel (2007) have convincingly argued, many (or most) apparent biological examples of downward causation (in particular cases of 'downward' regulation) seem to be adequately explainable by hybrid accounts appealing to intra-level causal interactions *between* constituents and inter-level constitutive relations. In those cases, an advocate of the constitutive interpretation of mereological supervenience could argue that the appeal to nested causation seems precisely to stem from an inadequate understanding of the role of configurations: the behaviour of the constituents appears to be influenced by the whole because the description focuses only on the internal properties of the constituents, by neglecting the relational ones.[24] In a word, there seems to be no clear case in the biological domain for which the appeal to nested causation is mandatory. Self-maintenance and closure are no exceptions in this respect.

The second strategy is epistemological, and consists in justifying nested causation by demonstrating that it would be impossible, *in principle*, to determine the behaviour of a system through a description of its configurational properties. On the basis of such negative result, the appeal to nested causation would be justified in epistemological terms, since there would be, in principle, no alternative description.[25] Yet, while arguments of 'inaccessibility' have already been formulated in physics (Silberstein and McGeever 1999), this is not the case in biology.[26] As a

[24] In the case of the wheel, for instance, one may say that if we describe a given molecule as a constituent of a wheel, we are already including in the description all constitutive and relational properties, which make it a constituent ('being in such and such position', 'having such and such interactions and links with neighbouring molecules' …), and which determine its behaviour under specific conditions. For instance, a force (i.e. gravity) applied to a part will generate a chain of causal interactions among the constituents which, because of their individual configurational properties, will behave in a specific way. We will then call the collective pattern the 'rolling movement of the wheel'. Each molecule of the wheel will move in a specific way because its configurational properties force it to do so, and a complete description of the configurational properties of the individual constituent will suffice to explain why it behaves as it does. The fact that the constituents collectively constitute a wheel, whose macroscopic behaviour can be described as a rolling movement, does not add anything to the explanation of the individual behaviour. There are indeed causal interactions here, but not inter-level causation.

[25] See Bich (2012) for an epistemological discussion of the relationship between emergence and downward causation.

[26] It should be noted, however, that the issue is currently being explored by several biologists and theoreticians. For instance, a relevant proposal in this direction has recently been developed by Soto, Sonnenschein and Miquel (Soto et al. 2008).

consequence, there seems to be, to date, no compelling epistemological arguments imposing to admit nested causation for biological systems.

The third strategy is heuristic. There are in fact many cases, especially in complex systems, in which the available description of the configurational properties is insufficient to adequately determine the behaviour of the whole system. In those cases, which are indeed widespread, it might be useful to appeal to the configuration as a whole *as if*, by virtue of its emergent properties, it were exerting nested causation on its constituents, so to provide a description capable of sufficiently determining the behaviour of the system. Since it is not committed to a theoretical non-constitutive interpretation of supervenience, the heuristic appeal to nested causation can be adopted as a pragmatic tool even within a constitutive interpretation of supervenience. Yet, such a heuristic use of nested causation would not point to specific features of the causal regime at work in biology (which is the object of this paper), but it would simply correspond to a scientific practice common to several scientific domains. In particular, as we mentioned above, the strategy can be adopted for self-maintaining and closed systems for which, mostly because of their internal complexity, complete descriptions of their organization are difficult to elaborate.

8 Conclusions

The main claim of this paper is that the organization of biological systems can be shown to realize a distinctive causal regime, that we labelled closure. Closure translates into contemporary terms the original Kantian idea, according to which living systems can be conceived as natural purposes, in which each part exists with respect to the other parts, such that the whole is able to self-maintain. In order for closure to be a legitimate scientific concept—and not just an epistemic shortcut—philosophical arguments must be provided in favour of its emergent and irreducible character with respect to the causal regimes at work in other classes of natural systems. To attain this objective, we developed a twofold argument.

On the one hand, we argued that closure is to be conceived as the mutual dependence among a set of constituents, each of them acting as a constraint. Constraints are configurations which, by virtue of the relations among their own constituents, possess emergent properties enabling them to exert distinctive causal powers on their surroundings, and specifically on thermodynamic processes and reactions. When a set of constraints realizes closure, the resulting organization constitutes a *higher-level emergent regime of causation*, possessing irreducible properties and causal powers. In particular, closed organizations are able to self-maintain as a whole (whereas none of the constitutive constraints can do it) which, in turn, enables them to generate biological *functions.*

On the other hand, we advocated the idea that a coherent defence of closure as an emergent and irreducible causal regime does not need to invoke nested causation. Closed organizations can be understood in terms of causal interactions among mutually dependent (sets of) constraints, without implying upward or downward causal actions between the whole and the parts. Biological emergence, accordingly,

is logically distinct from nested causation, and one can advocate the former without being committed with the latter.

Again we by no means want to exclude the possibility that biological organization might involve nested causation. As we discussed, various strategies could be adopted to adequately justify this idea, and promising explorations are currently underway. Nevertheless, we believe that these attempts are, as yet, incomplete, and do not offer compelling arguments in the biological domain. That is why we argued that biological organization can be shown to be emergent and irreducible *even though* nested causation is, by hypothesis, ruled out.

Acknowledgments The authors wish to thank Jon Umerez and Agustin Vicente for very valuable feedback on earlier versions of this paper. This work was funded by Ministerio de Ciencia y Innovación, Spain ('Juan de la Cierva' program to LB); Research Projects of the Spanish Government (FFU2009-12895-CO2-02 to AM and FFI2011-25665 to AM and LB); Research Projects of the Basque Government (IT 505-10 and IT 590-13 to AM and LB).

References

Alexander, S. (1920). *Space, time and deity*. London: Macmillan.

Bich, L. (2012). Complex emergence and the living organization: An epistemological framework for biology. *Synthese, 185*, 215–232.

Bitbol, M. (2007). Ontology, matter and emergence. *Phenomenology and the Cognitive Science, 6*, 293–307.

Block, N. (2003). Do causal powers drain away? *Philosophy and Phenomenological Research, 67*(1), 133–150.

Broad, C. D. (1925). *The mind and its place in nature*. London: Routledge and Kegan Paul Ltd.

Campbell, D. T. (1974). Downward causation in hierarchically organized biological systems. In F. J. Ayala & T. Dobzhansky (Eds.), *Studies in the philosophy of biology* (pp. 179–186). Berkeley and Los Angeles: University of California Press.

Campbell, R. J., & Bickhard, M. H. (2011). Physicalism, emergence and downward causation. *Axiomathes*, 21(1), 33–56. Quotations from the online version: http://www.lehigh.edu/~mhb0/physicalemergence.pdf.

Chandler, J. L. R., & Van De Vijver, G. (Eds.). (2000). *Closure: Emergent organizations and their dynamics* (Vol. 901). New York: Annals of the New York Academy of Science.

Chandrasekhar, S. (1961). *Hydrodynamic and hydromagnetic stability*. Oxford: Clarendon Press.

Cornish-Bowden, A., Cárdenas, M. L., Letelier, J.-C., & Soto Andrade, J. (2007). Beyond reductionism: Metabolic circularity as a guiding vision for a real biology of systems. *Proteomics, 7*, 839–845.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547–563.

Crutchfield, J. P. (1994). The calculi of emergence: Computations, dynamics, and induction. *Physica D: Nonlinear Phenomena, 75*, 11–54.

Emmeche, C., Køppe, S., & Stjernfelt, F. (2000). Levels, emergence, and three versions of downward causation. In P. B. Andersen, C. Emmeche, N. O. Finnemann, & P. V. Christensen (Eds.), *Downward causation* (pp. 13–34). Aarhus: Aarhus University Press.

Fell, D. (1997). *Understanding the control of metabolism*. London: Portland University Press.

Ganti, T. (1975). Organization of chemical reactions into dividing and metabolizing units: The chemotons. *BioSystems, 7*, 15–21.

Ganti, T. (2003). *The principles of life*. Oxford: Oxford University Press.

Glansdorff, P., & Prigogine, I. (1971). *Thermodynamics of structure, stability and fluctuations*. London: Wiley.

Hofmeyr, H.-J. S. (2007). The biochemical factory that autonomously fabricates itself: A systems biological view of the living cell. In F. Boogerd, F. J. Bruggerman, J.-H. S. Hofmeyr, & H. V. Westerhoff (Eds.), *Systems biology: Philosophical foundations* (pp. 217–242). Amsterdam: Elsevier.

Juarrero, A. (2009). Top-down causation and autonomy in complex systems. In N. Murphy, G. Ellis, & T. O'Connor (Eds.), *Downward causation and the neurobiology of free will* (pp. 83–102). Berlin: Springer.

Kant, E. [1781] (1987). *Kritik der Urteilskraft* (*Critique of Judgment*). Indianapolis: Hackett Publishing.

Kauffman, S. (2000). *Investigations*. Oxford: Oxford University Press.

Kim, J. (1993). *Supervenience and mind: Selected philosophical essays*. Cambridge: Cambridge University Press.

Kim, J. (1997). Explanation, prediction and reduction in emergentism. *Intellectica, 25*(2), 45–57.

Kim, J. (1998). *Mind in a physical world*. Cambridge: MIT Press.

Kim, J. (2003). Blocking causal drainage and other maintenance chores with mental causation. *Philosophy and Phenomenological Research, 67*(1), 151–176.

Kim, J. (2006). Emergence: Core Ideas and Issues. *Synthese, 151*(3), 547–559.

Kim, J. (2010). *Essays in the metaphysics of mind*. Oxford: Oxford University Press.

Laughlin, R., & Pines, D. (2000). The theory of everything. *Proceedings of the National Academy of Science of the United States of America, 97*, 28–31.

Laughlin, R., Pines, D., Schmalien, J., Stojkovic, B., & Wolynes, P. (2000). The middle way. *Proceedings of the National Academy of Science of the United States of America, 97*, 32–37.

Lloyd Morgan, C. (1923). *Emergent evolution*. London: Williams and Norgate.

Luisi, P.-L. (2006). *The emergence of life: From chemical origins to synthetic biology*. Cambridge: Cambridge University Press.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition. The realization of the living*. Dordrecht: Reidel Publishing.

Mayr, E. (2004). *What makes biology unique? Considerations on the autonomy of a scientific discipline*. Cambridge: Cambridge University Press.

McLaughlin, B. P. (1992). The rise and fall of British emergentism. In A. Beckermann, H. Flohr & J. Kim (Eds.), *Emergence or reduction? Essays on the prospects of nonreductive physicalism* (pp. 49–93). Berlin: Walter de Gruyter.

Mill, J. S. (1843). *A system of logic*. London: Parker.

Mossio, M. (2013). Closure. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho & H. Yokota (Eds.), *Encyclopedia of systems biology*. New York: Springer.

Mossio, M., & Moreno, A. (2010). Organizational closure in biological organisms. *History and Philosophy of the Life Sciences, 32*(2–3), 269–288.

Mossio, M., Saborido, C., & Moreno, A. (2009). An organizational account of biological functions. *British Journal for the Philosophy of Science, 60*, 813–841.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. New York: Wiley.

Pattee, H. H. (1972). Laws, constraints, symbols and languages. In C. H. Waddington (Ed.), *Towards a theoretical biology 4: Essays* (pp. 248–258). Edinburgh: Edinburgh University Press.

Pattee, H. H. (Ed.) (1973). *Hierarchy theory. The challenge of complex systems*. New York: Georges Braziller.

Piaget, J. (1967). *Biologie et Connaissance*. Paris: Gallimard.

Rosen, R. (1972). Some relational cell models: The metabolism-repair systems. In R. Rosen (Ed.), *Foundations of mathematical biology* (Vol. II, pp. 217–253). New York: Academic Press.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin and fabrication of life*. New York: Columbia University Press.

Ruiz-Mirazo, K. (2001). *Physical conditions for the appearance of autonomous systems with open-ended evolutionary capacities*. PhD Dissertation, University of the Basque Country.

Saborido, C., Mossio, M., & Moreno, A. (2011). Biological organization and cross-generation functions. *The British Journal for the Philosophy of Science, 62*, 583–606.

Salmon, W. C. (1998). *Causality and explanation*. Oxford: Oxford University Press.

Silberstein, M., & McGeever, J. (1999). The search for ontological emergence. *Philosophical Quarterly, 50*(195), 182–200.

Soto, A. M., Sonnenschein, C., & Miquel, P. A. (2008). On physicalism and downward causation in developmental and cancer biology. *Acta Biotheoretica, 56*, 257–274.

Sperry, R. W. (1969). A modified concept of consciousness. *Psychological Review, 76*(6), 532–536.

Stephan, A. (1992). Emergence—A systematic view on its historical facets. In A. Beckermann, H. Flohr, & J. Kim (Eds.), *Emergence or reduction? Essays on the prospects of nonreductive physicalism* (pp. 25–48). Berlin: Walter de Gruyter.

Teller, P. (1986). Relational holism and quantum mechanics. *The British Journal for the Philosophy of Science, 37*, 71–81.

Thompson, E. (2007). *Mind in life: biology, phenomenology, and the sciences of mind*. Cambridge: The Belknap Press of Harvard University Press.

Umerez, J., & Mossio, M. (2013). Constraint. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho & H. Yokota (Eds.), *Encyclopedia of systems biology*. New York: Springer.

Van Gulick, R. (2001). Reduction, emergence and other recent options on the mind/body problem: A philosophical overview. *Journal of Consciousness Studies, 8*(9–10), 1–34.

Varela, F. (1979). *Principles of biological autonomy*. New York: North Holland.

Varela, F., Maturana, H., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems, 5*, 187–196.

Vicente, A. (2011). Current physics and 'the physical'. *The British Journal for the Philosophy of Science, 62*(2), 393–416.

Weber, A., & Varela, F. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences, 1*, 97–125.